

[19] 中华人民共和国国家知识产权局

[51] Int. Cl⁷

H04Q 3/64

H04Q 3/52 H04B 10/12

H04B 10/24 H04B 10/08

[12] 发明专利申请公开说明书

[21] 申请号 01143804.5

[43] 公开日 2002 年 7 月 17 日

[11] 公开号 CN 1359241A

[22] 申请日 2001.12.13 [21] 申请号 01143804.5

[74] 专利代理机构 中国国际贸易促进委员会专利商标事务所

[30] 优先权

代理人 蒋世迅

[32] 2000.12.14 [33] EP [31] 00311184.6

[71] 申请人 朗迅科技公司

地址 美国新泽西州

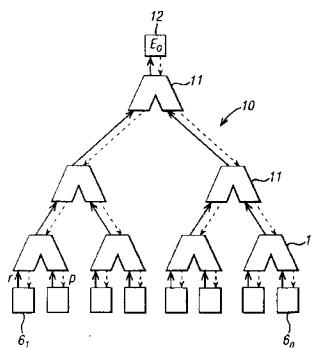
[72] 发明人 杰瑞恩·沃伦

权利要求书 3 页 说明书 10 页 附图页数 6 页

[54] 发明名称 用于分组交换机和无源光网络的分布式调度器

[57] 摘要

一种调度系统和方法,用于从输入端口($1_1 \cdots 1_i$)调度数据分组到输出端口($3_1 \cdots 3_n$),它包括虚拟输出队列($6_1 \cdots 6_n$),虚拟输出队列($6_1 \cdots 6_n$)安排成存储来自输入端口($1_1 \cdots 1_i$)的数据分组,其目的地是一个特定的输出端口($3_1 \cdots 3_n$)。调度系统包括有多个比较层的调度树(10),每个比较层安排成对比较从相关虚拟输出队列($6_1 \cdots 6_n$)并行接收的请求,以及在剩下单个请求之前,发送较高优先级请求到较高级比较层,单个请求指出被调度的虚拟输出队列($6_1 \cdots 6_n$)发送它的数据分组到相关的输出端口($3_1 \cdots 3_n$)。



ISSN 1 0 0 8 - 4 2 7 4

权 利 要 求 书

1. 一种从多个输入端口 ($1_1 \cdots 1_i$) 调度数据分组到至少一个输出端口 ($3_1 \cdots 3_o$) 的方法，该方法包括以下步骤：

在多个虚拟输出队列 ($6_1 \cdots 6_n$) 中存储数据分组，虚拟输出队列 ($6_1 \cdots 6_n$) 安排成存储来自多个输入端口 ($1_1 \cdots 1_i$) 中一个端口的数据分组，其目的地是该至少一个输出端口 ($3_1 \cdots 3_o$) 中特定的一个输出端口；和

调度多个虚拟输出队列 ($6_1 \cdots 6_n$)，其特征是，

调度多个虚拟输出队列 ($6_1 \cdots 6_n$) 的步骤包括以下步骤：

借助于调度树 (10)，调度与至少一个输出端口 ($3_1 \cdots 3_o$) 中一个端口相关的虚拟输出队列 ($6_1 \cdots 6_n$)，从而并行地调度与该至少一个输出端口 ($3_1 \cdots 3_o$) 中一个端口相关的虚拟输出队列 ($6_1 \cdots 6_n$)，

调度树 (10) 至少包括执行以下步骤的一个比较层：

成对比较从相关虚拟输出队列 ($6_1 \cdots 6_n$) 并行接收的请求；和

在剩下单个请求之前，发送较高优先级的请求到较高级比较层，单个请求指出被调度的虚拟输出队列 ($6_1 \cdots 6_n$) 发送它的数据分组到相关的输出端口 ($3_1 \cdots 3_o$)。

2. 按照权利要求 1 的方法，其中请求包括识别相关的虚拟输出队列 ($6_1 \cdots 6_n$)。

3. 按照权利要求 1 或 2 的方法，其中比较层还执行存储较高优先级请求的步骤，在从较高级接收到包括单个请求的准许之后，按照与较高优先级相关的存储请求，发送该准许到较低级比较层。

4. 按照权利要求 1, 2 或 3 的方法，其中成对比较请求步骤是对两个接收请求中的一个请求应用固定优先权。

5. 按照权利要求 1, 2 或 3 的方法，其中成对比较请求步骤是对两个接收请求中的每个请求应用交替优先权。

6. 按照权利要求 1, 2 或 3 的方法，其中请求包括优先级，以及成对比较请求步骤是比较各个优先级。

7. 按照以上权利要求之一的方法，其中比较层还执行以下步骤：

传输与较高优先级请求相关的数据分组到较高级比较层。

8. 一种调度系统，用于从多个输入端口（ $1_1 \cdots 1_i$ ）调度数据分组到至少一个输出端口（ $3_1 \cdots 3_o$ ），该系统包括虚拟输出队列（ $6_1 \cdots 6_n$ ），虚拟输出队列（ $6_1 \cdots 6_n$ ）安排成存储来自多个输入端口（ $1_1 \cdots 1_i$ ）中一个端口的数据分组，其目的地是该至少一个输出端口（ $3_1 \cdots 3_o$ ）中特定的一个输出端口，其特征是，

调度系统包括有多个比较层的调度树（10），每个比较层至少包括一个比较单元（11），比较单元（11）包括：两个输入门和一个输出门，最低级比较层中比较单元（11）的输入门连接到多个虚拟输出队列（ $6_1 \cdots 6_n$ ），比较层中两个比较单元（11）的输出门连接到较高级比较层中一个比较单元（11）的输入门，以及每个比较单元（11）安排成评定它的输入门接收到的各个请求并提供最高优先级请求到它的输出门。

9. 按照权利要求8的调度系统，其中请求包括识别相关的虚拟输出队列（ $6_1 \cdots 6_n$ ）。

10. 按照权利要求8或9的调度系统，其中至少一个比较单元（11）包括存储器装置，用于存储较高优先级的请求，且比较单元（11）还安排成从较高级比较层接收准许，该准许包括在最高级比较层中有最高优先级的请求，并按照与较高优先级相关的存储请求，发送该准许到较高级比较层中的比较单元（11）。

11. 按照权利要求8, 9或10的调度系统，其中该至少一个比较单元（11）安排成对它的两个输入门中一个输入门应用固定优先权。

12. 按照权利要求8, 9或10的调度系统，其中该至少一个比较单元（11）安排成对它的两个输入门应用交替优先权。

13. 按照权利要求8, 9或10的调度系统，其中请求包括优先级，以及该至少一个比较单元（11）安排成比较各个请求的优先级。

14. 按照权利要求8至13中一个的调度系统，其中该至少一个比较单元（11）包括：数据路径，用于传输与最高优先级请求相关的数据分组到较高级比较层。

15. 按照权利要求8至14中一个的调度系统，包括串联连接的多个

0.1.12.10

调度树 (10; 15, 16)。

16. 按照权利要求 8 至 14 之一的调度系统，包括与第一输出端口和第二输出端口相关的多个调度树 (10; 15, 16)，若相关的虚拟输出队列 ($6_1 \cdots 6_n$) 没有可用的第一输出端口，则调度系统安排成激活与第二输出端口相关的调度树 (10; 15, 16)。

说 明 书

用于分组交换机和无源光网络的分布式调度器

技术领域

本发明涉及用于分组交换机的调度器，具体涉及从多个输入端口调度数据分组到至少一个输出端口的方法，该方法包括以下步骤：在多个虚拟输出队列中存储数据分组，虚拟输出队列安排成存储来自多个输入端口中一个输入端口的数据分组，其目的地是该至少一个输出端口中特定的一个输出端口，以及调度多个虚拟输出队列。

背景技术

在太拉比特交换机和吉比特无源光网络（PON）中调度分组要求相当大的计算功率量。若必须部署优先级机构以管理不同服务质量（QoS）的业务，则问题将变得更加复杂。这种复杂性可以表示成系统中每个输出端口需要调度的输入队列总数，即，输入端口数目与服务等级数目的乘积。需要有这样一种算法，能够按照它们具体的优先级调度大量队列中的分组。在最新技术中，即，ASIC 或 FPGA，必须有效地执行这种算法。

C.Blondia, O.Casals 和 J.Garcia 的文章：‘A Cell Based MAC Protocol with Traffic Shaping and a Global FIFO Strategy’，Proceedings of the RACE Open Workshop on Broadband Access, Nijmegen, The Netherlands, June 1993, 公开一种媒体接入协议，它利用部署普通的先进先出（FIFO）缓冲器的请求/准许机构。每个网络终端（NT）通过请求广告它的带宽要求，包括 NT 中队列状态信息。利用带宽分配算法，媒体接入协议分配可用的带宽给各个 NT。借助于准许通知 NT 有关分配的带宽。这种用于 PON 的算法（具体地说，异步转移方式（ATM）PON）只寻址少量的队列（~64 个队列），不适合于吉比特容量的大系统（~1000 个队列）。此外，还要求连接 PON 到核心网的附加交换功能。

I.Elhanany, J Nir, D.Sadot 的文章：‘A Contention-Free Packet

Scheduling Scheme for Provision of Quality-of-Service in Tbit/sec WDM Networks' , Optical Networks Magazine, July 2000, 公开一种分组交换机的调度方案。提出的算法声称每个分组时隙周期约 $N^2 \log (N)$ 次操作, 其中 N 是输出端口或目的地数目 (该文章涉及 $N \times N$ 交换机)。利用循环过程, 包括每个输入端口的优先匹配方案以符合各种服务质量的要求, 这种方法采用顺序断言不同的输入端口。对于大量的队列, 这种方法仍然太慢。它还不能寻址 PON。

发明内容

本发明试图提供一种用于分组交换机和 PON 的调度器, 它能够按照它们特定优先等级调度大量队列中的数据分组。队列的数目等于输入端口的数目, 或在管理有不同服务质量要求的数据业务情况下, 队列的数目等于输入端口数目与服务等级 (或优先等级) 数目的乘积。

本发明提供一种按照上述前序部分的方法, 其中调度多个虚拟输出队列的步骤包括: 借助于调度树, 调度与至少一个输出端口中一个端口相关的虚拟输出队列, 从而并行地调度与该至少一个输出端口中一个端口相关的虚拟输出队列, 调度树至少包括一个比较层, 用于执行成对比较从相关虚拟输出队列并行接收的请求, 以及在剩下单个请求之前, 发送较高优先级请求到较高级比较层, 单个请求指出被调度的虚拟输出队列发送它的数据分组到相关的输出端口。

按照本发明的方法有以下的优点, 可以有效地调度非常大量的虚拟输出队列。本发明的调度方法只需要 $2 \log N$ 次操作, 其中 N 是虚拟输出队列的数目。通过级联式接入共享媒体和接入输出端口, 该方法不但可以有效地用于分组交换机, 还可以用于无源光网络。在相关的分组交换机或无源光网络中所有输出端口, 可以并行地执行本发明方法。

在本发明方法的一个实施例中, 请求包括识别相关的虚拟输出队列。它允许直接识别准许接入到某个输出端口的虚拟输出队列。

在另一个实施例中, 比较层还执行存储较高优先级请求的步骤, 在从较高级接收到包括单个请求的准许之后, 按照与较高优先级相关的存储请求, 发送该准许到较高级比较层。这个实施例可以简化分配机构, 避

免虚拟输出队列识别必须在调度树中传输。当然，这种反向路由调度方案要求总数为 $2^2 \log N$ 次操作，这种方案操作次数仍然少于现有技术调度方案的操作次数。

成对比较请求可以给两个接收请求中的一个请求应用固定优先权，允许非常简单地实施有不同程度固定端口/服务等级组合的方法。作为另一种方案，可以给两个接收请求中的每个请求应用交替优先权，最终效果是，没有给被调度的虚拟输出队列中任何队列授予优先权。在另一种方案中，请求包括优先级，以及成对比较请求的步骤应用各个优先级的比较。这个实施例可以灵活地对虚拟输出队列进行编程，其代价是更加复杂的比较逻辑电路。增加的复杂性要求约 $2 \log N_c$ 个附加的逻辑电路，其中 N_c 是可用的优先级数目。

在另一个本发明的实施例中，比较层还执行步骤：传输与较高优先级请求相关的数据分组到较高级比较层。利用这个实施例，在与建立调度机构的同时，可以在输入端口与输出端口之间建立数据分组的路径。

在另一方面，本发明涉一种调度系统，用于从多个输入端口调度数据分组到至少一个输出端口，该系统包括虚拟输出队列，虚拟输出队列安排成存储来自多个输入端口中一个输入端口的数据分组，其目的地是该至少一个输出端口中一个特定输出端口。调度系统包括有多个比较层的调度树，每个比较层至少包括一个比较单元，比较单元包括两个输入门和一个输出门，最低级比较层中比较单元的输入门连接到多个虚拟输出队列，比较层中两个比较单元的输出门连接到较高级比较层中一个比较单元的输入门，以及每个比较单元安排成评定它的输入门接收到的请求，并提供最高优先级请求到它的输出门。利用组合的逻辑电路可以容易地实现本发明的调度系统，而且可以在单个 ASIC 中与交叉连接交换机集成。此外，调度树可以按照分布和可扩充的方式实现，例如，在与若干个输入端口相关和每个输出端口相关的两层区域中，且位于对应的线路卡上。

为了能够直接识别准许接入输出端口的虚拟输出队列，该请求最好包括识别相关的虚拟输出队列。

在本发明调度系统的一个实施例中，该至少一个比较单元包括存储较高优先级请求的存储器装置，以及比较单元还安排成接收来自较高级比较层的准许，该准许包括最高级比较层中有最高优先级的请求，并按照与较高优先级相关的存储请求，发送该准许到较低级比较层中的比较单元。

该至少一个比较单元可以安排成给它两个输入门中的一个输入门应用固定优先权或交替优先权。

或者，该请求包括优先级，以及至少一个比较单元安排成比较各个请求的优先级。

在另一个实施例中，该至少一个比较单元包括数据路径，用于传输与最高优先级请求相关的数据分组到较高级比较层。

为了能够从若干个输入端口调度接入到若干个输出端口中一个输出端口，可以串联地连接多个调度树。这可应用在要求调度接入共享媒体（共享光纤）和接入输出端口的无源光网络中。

在本发明调度系统的另一个实施例中，调度系统包括与第一输出端口和第二输出端口相关的多个调度树。若相关的虚拟输出队列没有可用的第一输出端口，则调度系统安排成激活与第二输出端口相关的调度树。这个实施例允许无连接调度操作或保护交换操作，以及可以容易地利用附加的逻辑电路实现。

附图说明

利用若干个典型的实施例并参照附图详细地说明本发明，其中：

图 1a 表示连接输入端口到输出端口的交换机基本配置；

图 1b 表示无源光网络的基本配置；

图 2 表示以示意图形式调度有某种服务质量的数据业务；

图 3 表示按照本发明的调度树示意图；

图 4a-c 表示本发明多级调度装置的实施例示意图；

图 5a 表示有反向准许路由的固定优先权比较器的示意图；

图 5b 表示按照本发明调度器的集成路径部分的示意图；

图 6 表示在分组交换机调度器中实现本发明的示意图；和

图 7 表示在无源光网络调度器中实现本发明的示意图。

具体实施方式

图 1a 表示应用调度器交换输入数据流到特定输出数据流的示意图。

交换机 2 连接 N_i 个输入端口 $1_1 \cdots 1_i$ 到 N_o 个输出端口 $3_1 \cdots 3_o$ 。借助于交换机 2，每个输入端口 $1_1 \cdots 1_i$ 可以连接到任何的输出端口 $3_1 \cdots 3_o$ 。为了避免来自不同输入端口 $1_1 \cdots 1_i$ 的数据发生冲突，就需要调度功能以调整接入到每个输出端口 $3_1 \cdots 3_o$ 。在无源光网络 (PON) 中也需要类似的功能，如图 1b 所示，在交换机 2 附近，输入端口 $1_1 \cdots 1_i$ 连接到 N_g 个共享媒体 $4_1 \cdots 4_g$ 。在 PON 的情况下，调度功能不但必须调整接入到输出端口 $3_1 \cdots 3_o$ ，而且还接入到共享媒体 $4_1 \cdots 4_g$ 。在上述情况下，交换机 2 考虑成单级无阻塞的，即，输入端口 $1_1 \cdots 1_i$ 直接地连接到被调度的输出端口 $3_1 \cdots 3_o$ 。在 PON 的情况下，可以实现两个单独的调度器，一个调度器用于接入到共享媒体 $4_1 \cdots 4_g$ ，而另一个调度器用于接入到输出端口 $3_1 \cdots 3_o$ 。然而，这要求在交换机 2 的输入端有附加的输入队列，从而降低了端到端的性能。

为了避免在输出端口 $3_1 \cdots 3_o$ 的数据分组发生冲突，要对数据分组进行排队。排队数据分组可以有多种方法。数据分组可以存储在该分组要到达的输出端口 $3_1 \cdots 3_o$ ，但是，这意味着所有提供的数据分组必须首先通过交换机 2 传输。对于大量的端口数目，这就要求不切实际的内部速度。另一种可能性是集中排队，然而它要求大量复杂的电路，这在有大量输入端口和输出端口的情况下是很难实现的。还有一种可能性是在输入端口 $1_1 \cdots 1_i$ 对数据分组进行排队。这种方法的缺点是，寻址可用输出端口 $3_1 \cdots 3_o$ 的数据分组可能被相同队列中正在等待另一个输出端口 $3_1 \cdots 3_o$ 的数据分组所阻塞（称之为排头(head of line)阻塞）。可以采用称之为虚拟输出排队 (VOQ) 的方法克服这个缺点，其中数据存储在一个输出端口 $3_1 \cdots 3_o$ 专用的单独输入队列中，导致该系统中有 $N_i N_o$ 个被调度的虚拟端口。

当不同优先级的数据分组从输入端口 $1_1 \cdots 1_i$ （对应于不同的服务质量 QoS）传输时，调度功能还应当考虑到这些优先级。为了避免较高优先级数据分组被较低优先级数据分组所阻塞，可以扩充虚拟输出排队机构以包括 N_c 个优先等级，从而使被调度的队列数目等于 $N_i N_o N_c$ 。调度不同

队列优先级要求公平的权重。

图 2 表示调度有不同优先级的数据分组，从一个特定输入端口 1 调度数据业务到一个特定输出端口 3。分类器 5 评定输入的数据分组，并按照优先等级把它分配到多个虚拟输入队列 $6_1 \cdots 6_c$ 中的一个队列。各个队列 $6_1 \cdots 6_c$ 输入到调度器 7 以接入到输出端口 3，包括评定队列 $6_1 \cdots 6_c$ 的优先等级。这种调度操作是对 $N_i N_c$ 个虚拟输出队列中的每个队列进行的。

调度器 7 完成的调度过程可以分成三个阶段。首先，调度器 7 查询队列 6（对于有 N_i 个输入端口， N_o 个输出端口和 N_c 个优先等级的系统，队列总数为 $N_i N_o N_c$ ）的接入请求。其次，基于具体的优先级，调度器 7 确定哪个队列 6 被准许接入到输出端口 $3_1 \cdots 3_o$ 。最后，调度器 7 通知被准许的队列 6 发送相关的数据分组到被请求的输出端口 3。

这些步骤对于某个数据分组交换机 2 中的每个分组是反复进行的，因此，整个过程需要持续的时间小于相关的时隙 T_s 。在近代的数据分组网络中，时隙 T_s 为微秒的数量级。由于当前处理器工作的循环时间是几个纳秒的数量级，这就允许每个时隙中约 100 次操作以调度所有的虚拟输出队列。调度操作中大量消耗的时间是评定。已知的调度技术利用查询或循环方式，但是，这些技术对于大量队列来说是太慢了。利用分级方法把队列请求评定和加权分解成图 3 所示的二进制树 10，本发明可以对大量的队列进行调度操作。

在图 3 中，画出单个调度树 10，用于确定多个虚拟输出队列 $6_1 \cdots 6_n$ 中哪个队列可以接入到某个输出端口 3。因此，所示的调度树 10 是每个输出端口 $3_1 \cdots 3_o$ 所需要的。

调度树 10 包括若干个相同的比较单元 11，例如，它可以是用组合逻辑电路实现的（见以下）。每个比较单元 11 包括两个输入端和一个输出端。比较单元 11 以树形的方式互相连接，即，较高级比较单元 11 的一个输入端连接到较低级比较单元 11 的输出端。在最低级上，比较单元 11 的输入端连接到虚拟输出队列 $6_1 \cdots 6_n$ 。在每个时隙 T_s ，虚拟输出队列 $6_1 \cdots 6_n$ 在比较单元 11 的一个输入端指出它们的请求 r 。占优势的请求传送到较高级，直至最高级，在 $2 \log N$ 次门操作之后，出现特定输出端口 3 的

准许请求。在最高级比较单元 11 的输出端，仅剩下特定输出端口 3 的一个请求，因此，该请求可以直接分配到相关的虚拟输出队列 $6_1 \cdots 6_n$ 。

通过选取合适的电路，比较单元 11 可以实现并行的工作，因此，只需要² $\log N$ 个步骤（其中 N 是每个输出端口需要调度的（虚拟）端口数目，即，在 N_i 个输入端口和 N_c 个服务等级的情况下， $N=N_iN_c$ ）以确定哪些虚拟输出队列 $6_1 \cdots 6_n$ 准许接入到特定输出端口 3。

例如，基于比较单元 11 中一个输入端占有优先权的固定优先权，比较单元 11 可以在任何级确定哪个请求 r 占优势。在图 3 所示的实施例中，树入口的优先级在左侧方向或右侧方向增大。这就简化了实现有不同程度固定端口和服务等级组合的系统。

或者，比较单元 11 可以安排成对接收到的请求给出交替优先权，即，在每次给一个输入端口授予准许 p 之后，优先权交换到另一个输入端口，这相当于没有优先权。

在另一种方案中，在比较单元 11 中可以实现加权优先权。于是，每个虚拟输出队列 $6_1 \cdots 6_n$ 的请求应当包括优先等级。在此情况下，灵活编程虚拟输出队列是可能的，但是它要求更复杂的逻辑电路（约为² $\log N_c$ 倍）以实现比较单元 11。

可以避免虚拟输出队列 $6_1 \cdots 6_n$ 的标识通过调度树 10 传输和简化分配机构。这可以把比较单元 11 安排成存储相关的中间结果来完成。按照这种方法，可以反向路由最高级比较单元 11 中输出端接收到的准许 p 通过调度树 10 到右侧虚拟输出队列 $6_1 \cdots 6_n$ 。这在图 3 中是用虚箭头指出的。

还可以安排调度树 10 包括有比较单元 11 的集成路径部分电路，从而启动调度器 7 在虚拟输出队列 $6_1 \cdots 6_n$ 与输出端口 $3_1 \cdots 3_o$ 之间建立数据分组的路径。

在此之前描述的调度树 10 允许调度数据分组到一个输出端口 3。对于分组交换机 2 中每个输出端口 $3_1 \cdots 3_o$ ，由于调度树 10 可以并行地工作，所以调度全部($N_iN_oN_c$ 个)虚拟输出队列 $6_1 \cdots 6_n$ 所需的步骤总数也等于² $\log N_iN_oN_c$ 。

通过按序放入两个调度树，本发明还可以控制接入到无源光网络中的

共享媒体 $4_1 \cdots 4_g$ 。图 4a 以简化的形式表示无源光网络中单个队列 6 的多级调度装置。首先，利用第一调度树 15 调度输入队列 $1_1 \cdots 1_i$ 到共享媒体 $4_1 \cdots 4_g$ ，其中实现反向路由安排，然后，利用第二调度树 16 调度共享媒体 $4_1 \cdots 4_g$ 到输出端口 $3_1 \cdots 3_o$ 。因此，两个调度树 15 和 16 互连可以完成 AND 功能。

图 4b 表示利用第一调度树 15 和第二调度树 16 的另一个可行实施例，其中第一调度树 15 得到的准许 p_1 被反转并输入到 AND 门 17，AND 门 17 的第二输入端连接到一个虚拟输出队列 6，而 AND 门 17 的输出端连接到第二调度树 16。实际上，这实现了接入到某个输出端口的逻辑 OR 功能，因此可以调度无连接方式的数据分组。所有虚拟输出队列 6 的第一调度树 15 应当首先输出准许 p_1 ，然后，准许 p_1 应当被 AND 门 17 处理并输入到第二调度树 16 以避免双重或未请求的分配。

图 4c 表示利用第一调度树 15 和第二调度树 16 的另一个可行实施例。来自虚拟输出队列 6 的请求输入到第一调度树 15 和第二调度树 16。从第一调度树 15 或从第二调度树 16 得到准许 p_1 。实际上，这也实现了接入到某个输出端口的逻辑 OR 功能，但现在仅允许支持保护交换。

图 5a 表示有组合逻辑电路的调度树 10 中典型比较单元 11 装置的简单方框图。比较单元 11 具有固定优先权和准许 p 的反向路由。第一输入端 20 和第二输入端 21 连接到 OR 单元 28。取决于第一输入端 20 或第二输入端 21 上的请求 r 输入值（高逻辑值或低逻辑值），比较单元 11 的输出（OR 单元 28 的输出）反映这个特定比较单元 11 是否发送请求到较高级比较单元 11。来自较高级比较单元 11，在比较单元 11 的准许输入端 23 接收到准许值。这个准许输入端 23 连接到第一 AND 门 26 和第二 AND 门 27 的输入端。当这个特定比较单元 11 从较高级比较单元 11 接收到准许 p 时，第一 AND 门 26 和第二 AND 门 27 输出它们各自第二输入端上的逻辑值。AND 门 26 的第二输入端连接到第一请求输入端 20，而 AND 门 27 的第二输入端连接到第一请求输入端 20 的反转值。AND 门 26 和 27 的输出端分别给出第一输入端 20 和第二输入端 21 的准许（逻辑值）。实际上，这给第一输入端 20 提供固定优先权。

图 5b 表示可以集成在比较单元 11 中的路径部分电路的典型实施例，

该电路在图 5a 组合逻辑电路附近。路径部分电路包括：第一启动门 33 和第二启动门 34。第一启动门 33 和第二启动门 34 的输入端分别接收与第一输入端 20 和第二输入端 21 相关的数据分组 30 和 31。第一启动门 33 和第二启动门 34 的输出端连接并形成路径部分输出端 32。第一启动门 33 和第二启动门 34 是受门输入端的控制，这两个门输入端接收图 5a 组合逻辑电路的准许输出 24 和 25。图 5a 和图 5b 的组合提供了实现比较单元 11 的电路单元，它自动地在输入端口 1 与输出端口 3 之间建立连接。这个电路单元可以构造一个完全的交换机 2。本领域专业人员清楚地知道，在比较单元 11 中，选通单元和锁存单元是提供合适定时和同步所必需的单元。

图 6 表示在分组交换机中应用本发明的示意图。在此情况下，调度树 10 可以分布在输入节点与输出节点之间，可能是通过多个级和底板级联以支持相关的物理尺寸。在图 6 所示的实施例中，16 x 4 交换机 2 的分组交换调度器 40 被分成三个部分，输入级 41，底板 42 和输出级 43。

输入级 41 分成连接到管理模块 44 的四个输入模块 $45_1 \cdots 45_4$ 。此外，输出级 43 分成四个输出模块 $49_1 \cdots 49_4$ 。每个输入模块 45_k 有四个输入队列 $46_{k1} \cdots 46_{k4}$ 和相关的警管控制器 $47_{k1} \cdots 47_{k4}$ ，输入队列用于从相关的输入端 $1_1 \cdots 1_4$ 接收数据分组，而警管控制器在从调度树 10 接收到准许 p 之后，允许输入队列 $46_{k1} \cdots 46_{k4}$ 发送它的数据分组。调度树 10 在与四个输出端口中一个端口相关的每个输入模块 45_k 中分成四个输入调度树 $48_{k1} \cdots 48_{k4}$ 和每个输出模块 $49_1 \cdots 49_4$ 中一个输出调度树 $50_1 \cdots 50_4$ 。输入调度树 $48_{k1} \cdots 48_{k4}$ 的输出端连接到底板 42，底板 42 连接输入调度树 $48_{k1} \cdots 48_{k4}$ 的输出端到相关的输出调度树 $50_1 \cdots 50_4$ 。管理模块 44 控制警管功能以及其他定时和同步功能。也可以用分布的方式实现警管功能，但应当总是位于输入队列 $46_{k1} \cdots 46_{k4}$ 与输入调度树 $48_{k1} \cdots 48_{k4}$ 之间。

在图 6 所示的实施例中，可以存在特殊的规定，例如，允许最高级请求直接环回到输入队列 46 的请求中或允许保护交换。

图 7 表示无源光网络 (PON) 中调度器 70 装置的示意图。图 6 所示分组交换调度器 40 与图 7 所示 PON 调度器 70 之间的主要差别是，PON 调度器 70 不是直接地连接到远程光网络单元 (ONU) 中的队列，中间隔着共享媒体 $4_1 \cdots 4_g$ 的附加评定级 (见图 1b 和以上的描述)。此外，PON

调度器 70 一般不配备图 5b 所示的集成路径部分，而是控制接入到一个或多个共享媒体 $4_1 \cdots 4_g$ ，共享媒体 $4_1 \cdots 4_g$ 连接到输出端口 $3_1 \cdots 3_n$ 。借助于分开的（光）交叉连接。图 7 表示 PON 调度器 70 也分成输入级 51，底板 52 和输出级 53。输出级 53 和底板 52 的功能和结构分别类似于分组交换调度器 40 中输出级 43 和底板 42 的功能和结构。然而，对于每个输入模块 $55_1 \cdots 55_3$ ，输入级 51 包括附加的共享媒体级和输入/输出（I/O）底板 $61_1 \cdots 61_3$ ，用于接收远程请求和发送远程准许到 PON 中 ONU。

为了调度驻留在远程光网络单元中的队列，调度器 70 必须提供通信装置以支持调度过程中的评定和分配阶段。虽然是可能的，但是把调度树 10 分布到 PON 中是切实际的，因为该领域中没有工作在链路级上的设备。在给出的实施例中，这是通过在中心位置引入队列代理 56 解决的。在图 7 中，作为例子画出每个输入模块 55_k 的两个队列代理 $56_{k1} \cdots 56_{k2}$ 。队列代理 $56_{k1} \cdots 56_{k2}$ 通常是给 ONU 建立的，这些 ONU 要求给它们的数据业务提供某个固定带宽。为了避免高服务质量（高优先级）的确定性数据业务的长时间往返行程延迟，这些队列代理 $56_{k1} \cdots 56_{k2}$ 可以独立地产生请求以支持远程队列。突发性业务的处理类似于图 6 中所示分组交换机 40 中的数据分组，即，发送这类数据业务的 ONU 必须提出请求给 PON 调度器 70。从输入调度树 $58_{k1} \cdots 58_{k4}$ 接收到的准许经 I/O 底板 $61_1 \cdots 61_3$ 送回到远程 ONU。

此外，必须配备 PON 调度器 70 以处理 PON 的维护功能。它是受 OAM 和测距块 62 的控制，特别是在要求测距操作使新 ONU 初始化的情况下。为了这些目的，可以配置有最高优先级的专用队列，因此，每当测距功能需要寂静周期时，它可以要求随后的时隙。

利用图 5a 中展示和描述的固定优先权方案，可以相当简单地实现队列代理 $56_{k1} \cdots 56_{k2}$ 。在此情况下，来自队列代理 $56_{k1} \cdots 56_{k2}$ 的请求 r 仅仅涉及单个逻辑值（‘0’ 或 ‘1’），利用指出未完成请求的普通计数器，可以实现本地排队功能。

说 明 书 附 图

图 1a

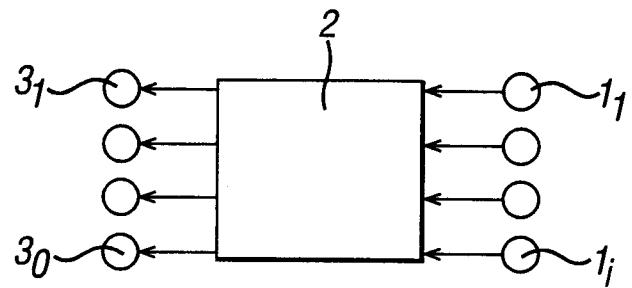


图 1b

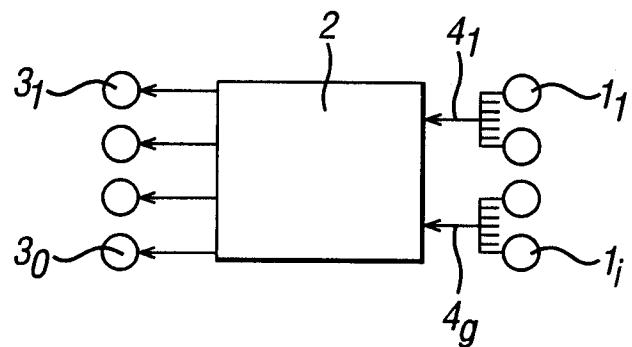


图 2

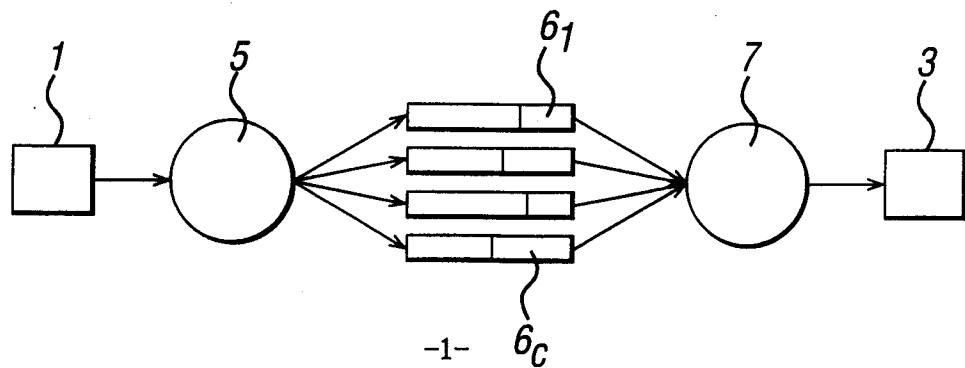
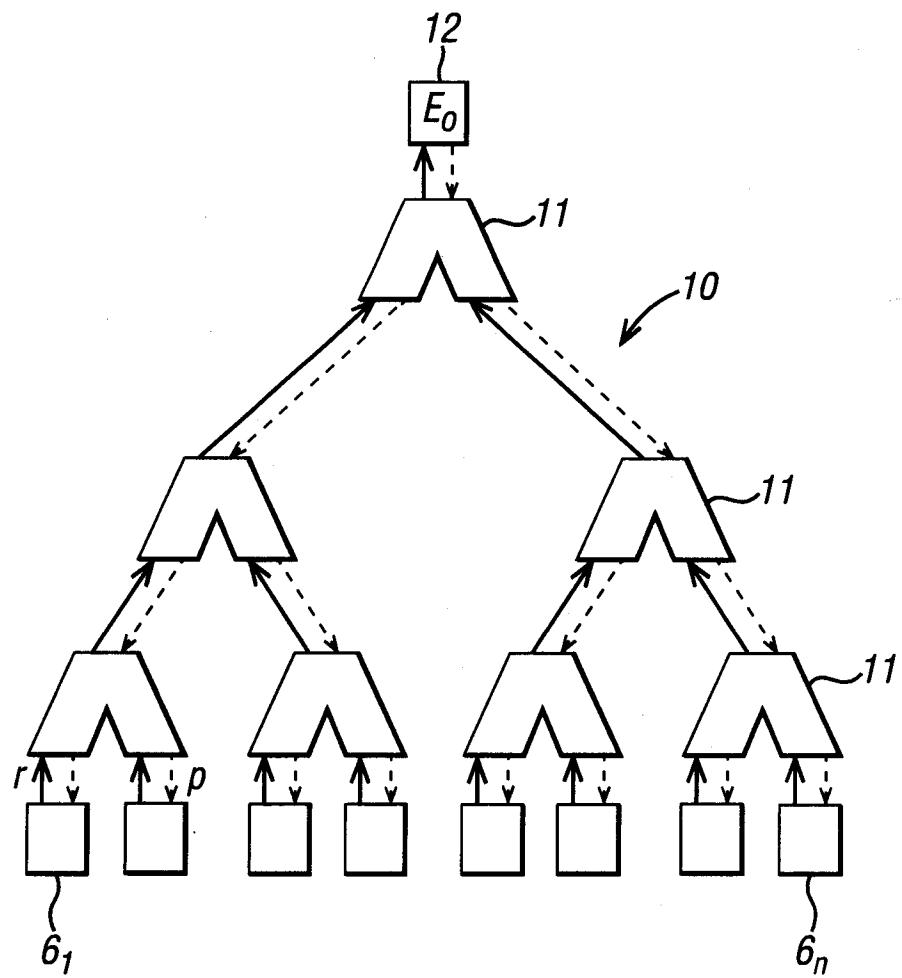
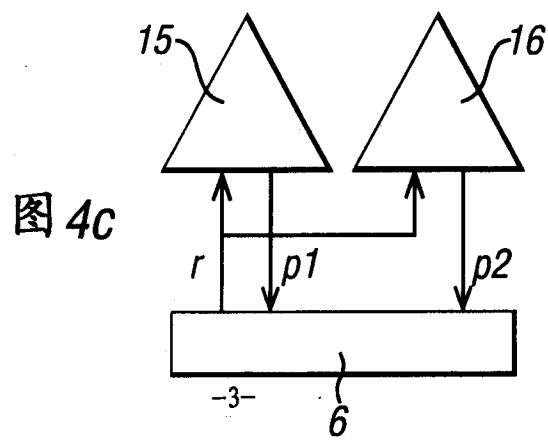
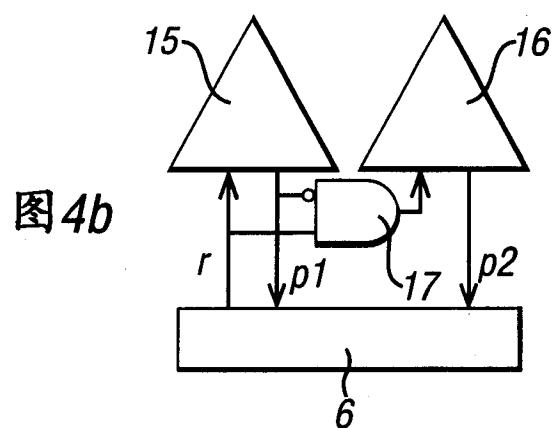
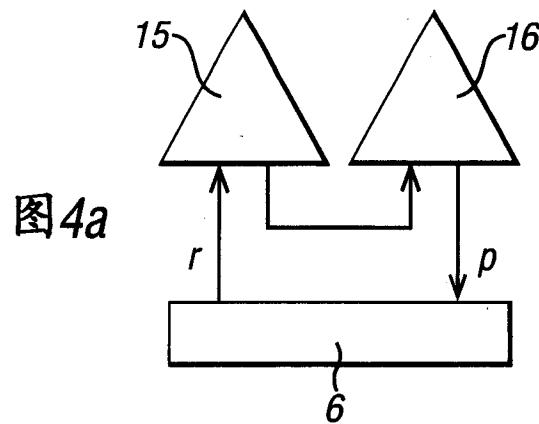


图3





00111213

图 5a

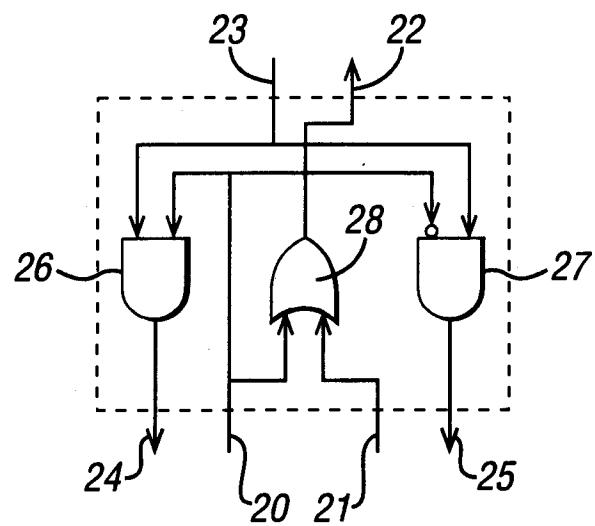


图 5b

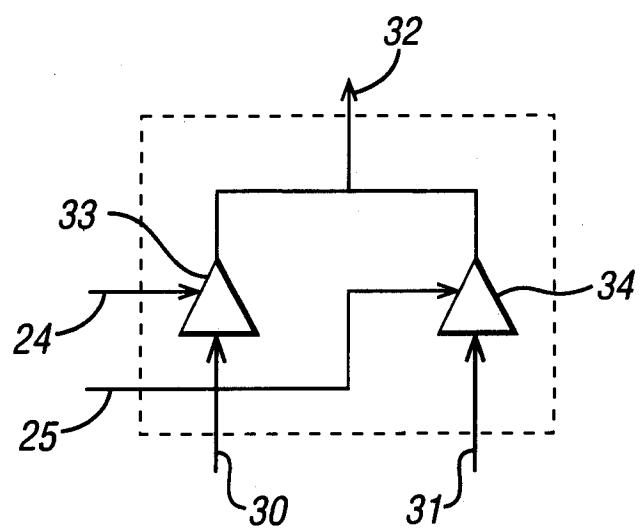


图6

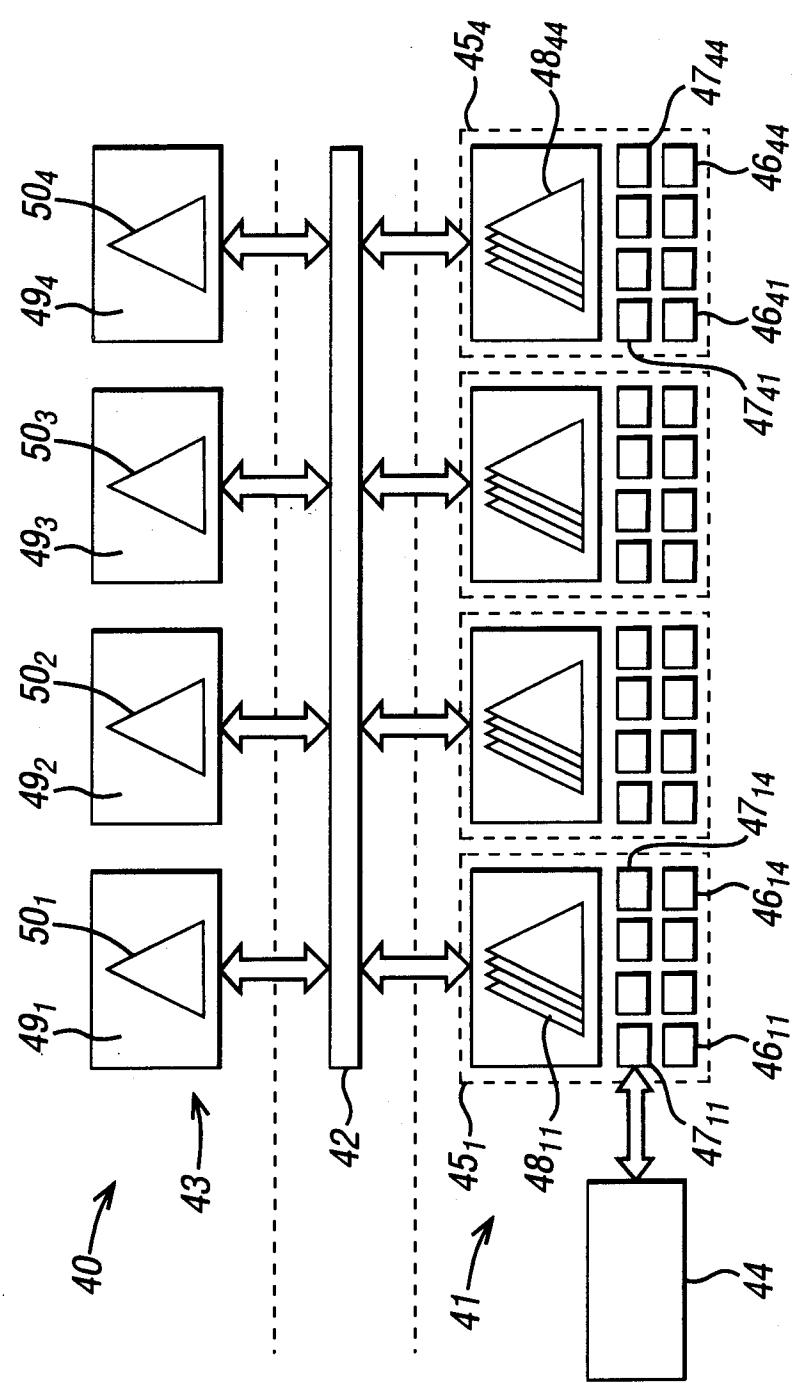


图7

